DOCUMENT RESUME

ED 086 192 IR 000 039

AUTHOR Hays, David G.
TITLE Cognitive Networks and Abstract Terminology.
PUB DATE Oct 73
NOTE 17p.; Paper presented at the Conference on Computers in Clinical Medicine (Buffalo, New York, October 29-31, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Codification; Computational Linguistics; *Information Needs; Information Science; *Information Storage; Information Theory; *Linguistic Theory; *Medical Vocabulary; *Medicine; Semantics
IDENTIFIERS SNOMed System

ABSTRACT

The design of codes and formats for information storage must be guided by the requirements of the processes in which the stored information is to be used. In the case of medical data, a code must be designed to facilitate several different processes, including case management, administration, evaluation and research. Design of such a system can be facilitated by an examination of the theory, in which knowledge is represented by a formal model consisting of nodes and arcs connecting the nodes. Arcs are of five types: paradigmatic, syntagmatic, discursive, attitudinal and metalingual. Metalingual arcs, which allow for shifts in level of abstraction, provide the only natural means of passage among the several languages of medicine. The technical problems raised by these shifts have been neglected in the philosophy of science, in linguistics and in information retrieval. Any cognitive network needs to record paradigmatic, syntagmatic, discursive and metalingual arcs; attitudinal arcs are also sometimes needed. Medical record files must be able to keep all types of linkages to provide accurate histories, and medical authority files need them to record the structure of the science. The SNOMed system provides some, but not all, of these linkages. (Author/SL)

FILMED FROM BEST AVAILABLE COPY

ED 026192

COGNITIVE NETWORKS AND ABSTRACT TERMINOLOGY

David G. Hays

Department of Linguistics

State University of New York, Buffalo

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Information Storage and Information Processing

The design of codes and formats for information storage must be guided by the requirements of the processes in which the stored information is to be used. However obvious this point may seem to others, I had to learn it from a colleague, Martin Kay; and it did not seem obvious to me for a long time. A general theory of information processing would summarize the requirements of all, or many, processes, and could thus serve the designers of codes and formats; but such a theory does not yet exist in adequate detail, and would in any case have to be specialized in applications. General theories always say less than special theories, and designers need guidance in numerous small matters.

The problem at hand is the storage of medical data. In particular, a code is proposed; it is to be detailed enough to record case histories and research. The code must surely serve at least the processes stated in the following paragraphs, although the list is not exhaustive.

Based on a lecture at the Fifth Buffalo Conference on Computers in Clinical Medicine, October 29-31, 1973. The paper benefits at several points from the Conference discussion. Submitted for publication in the Journal of Clinical Computing.

IR 000 039

Encoding. Physicians, nurses, technical personnel, or information specialists must register selected information in accordance with the code.

Case management. The code is intended to be useful to the physicians and staff with medical responsibility for a case.

Administration. The code is to provide a basis for the determination and allocation of charges, and justification of them.

Assessment. The code is intended to provide a basis for review of practice in a hospital or community.

Research. The files stored in the code are to serve analysts who wish to test and improve any part of the theory of medicine and health.

Modification. The code will have to be modified or replaced as the theory on which it is based grows and changes; modification, when feasible, is less expensive than replacement.

This table of processes is forbidding, and leads to such a long list of requirements as to discourage some observers. Nevertheless, the history of the field shows an active demand for codes that can serve all these and other processes, and encourages the designers to go on with their task. My purpose is to strengthen their hand. My method is to present some elements of a theory of knowledge and illustrate its application to their problem. Two

outcomes are foreseeable: By systematic application of this or a similar theory, the internal structure of the code can perhaps be improved slightly. And ancillary tools can be developed to facilitate the six processes listed above--among others.

Cognitive Networks

In artificial intelligence and increasingly in psychology, knowledge is represented by a formal model consisting of nodes and arcs connecting the nodes. The version I present here is the work of myself and students; but it differs only in small--but we think significant--details from numerous other versions in the literature (see the references at the end of the paper for citations).

Nodes are of four types: entities, events, properties, and modalities.

Arcs are of five types: paradigmatic, syntagmatic, discursive, attitudinal, and metalingual.

Assume that inflammation, hepatitis, and viral hepatitis are entities. Their paradigmatic relationships can be represented by noting that hepatitis is a variety of inflammation, and viral hepatitis is a variety of hepatitis. Paradigmatic relations are shown in SNOMed by the structure of the code; all of the codes in a field that have identical characters in the first several positions are related.

A second paradigmatic relationship is that of part to whole. Thus the tonsil is a part of the pharynx, but it is

also part of the lymphatic structure. As I understand the SNOMed system, a choice must be made by the designers in each case where alternative paradigmatic relationships appear. The structural decision to make such choices is the result of putting paradigmatic structure into the code, and wishing to avoid redundancy. The structural decision will restrict users of the code, I think unduly, and force the designers of process algorithms to introduce systematically searches for elements that do not appear in all their paradigmatic positions. If the tonsil, according to the code, belongs to the pharynx, students of the lymphatic system will be forced to take extra pains to find it; and if the code makes it part of the lymphatic system, the student of the pharynx will have to search to find it. Two ways out are apparent; either the code is made redundant, or a tool is published listing all such suppressed relationships--and there will no doubt be many.

Syntagmatic relationships give internal structure to propositions. One kind is the application of a property to an entity. If viral is a property, then viral applies to hepatitis; that is, hepatitis is sometimes characterized by the property of viral etiology. Also, on occasion, the property few applies to erythrocytes, or the property soft applies to bone.

Another basic set of syntagmatic relationships holds between entities and events; the entities are said to

participate in events as agents, datives, instruments, and patients. According to this theory, agents and datives are basically animate, but instruments and patients are not necessarily so. Agents and instruments are causally involved in the event, but datives and patients are not. Thus the abstract representation of "The nurse administered the drug to the patient with a hypodermic" takes the administration as the event. Nurse and patient are animate; nurse and hypodermic are causal; drug is neither. Hence nurse is agent, hypodermic is instrument, patient is dative; and drug is patient. (The fact that 'patient' appears in both the language of medicine and the language of cognitive networks is no more than an unfortunate etymological accident.)

SNOMed appears to have little capacity for representation of syntagmatic relationships, either explicitly or implicitly. I do not mean that it has none; but the designers have evidently not considered the problem worth extended labor. Syntagmatic relationships are often neglected except by linguists who, specializing in them, take them very seriously and may exaggerate their importance.

Discursive arcs represent spatial, temporal, and causal relationships between propositions. Through discursive links, the nurse's hypodermic administration of a certain drug to a certain patient can be fixed in time, in place--relative to either the hospital or the anatomy of the patient--and in causality, since the effect of the drug

would often be part of the record. Technically, a proposition consists of an event and participating entities, or a property and the entity to which it applies. The head of the proposition, which is the event or property, is linked to a modality; and the discursive arcs are attached to the modality. The reason for this circuitous linkage is that the entire proposition, not just its head, is placed in the framework of space, time, and causality; the modal node can be understood to represent the whole proposition.

Discursive relationships are evidently implicit in the structure of SNOMed (see Côté's contribution). The topographic field encodes space. The morphology field encodes an event. The etiological field encodes another event, causally related to the morphological event. The relationships of the other fields are more complex, I think.

Attitudinal arcs connect entities, but only those capable of holding attitudes, to propositions. Some types of attitudinal arcs are belief, desire, purpose, and doubt. Such relationships have a selfevident place in clinical records. A diagnosis or prescription must be ascribed to a physician who is willing to take responsibility for it in the future. SNOMed makes no provision for the encoding of these relationships, but presumably the format of a record will include the identification of the person to whom the content is attributed.

Abstract Terminology

Metalingual arcs permit a shift in level of abstraction. Every science relies on such shifts, and medicine, which has to be regarded as a collection of sciences, makes them in a strikingly fluid and frequent manner. The technical problems raised by these shifts are neglected in the philosophy of science, in linguistics, in information retrieval at large, and therefore not surprisingly in SNOMed.

Medicine speaks numerous languages. One way of grouping them is to speak of the clinical picture (history and physical examination), the laboratory findings (chemistry, radiology, etc.), and the pathophysiology (direct examination of tissues). These languages can be translated; indeed, the practice of clinical medicine is based on their intertranslatability. But the translation is not trivial. It is problematic, and depends on the state of science at the moment. Hence it is not to be performed casually, and certainly not implicitly. SNOMed should permit its users to record their data in all three languages, not forcing them to translate before encoding (even if some users choose to do just that). And it should permit the user to connect the levels according to known theory. The difficult question is how the connections are to be represented; the metalingual arc is my answer to that question.

A metalingual definition is any portion of a cognitive network, with a modality as head, linked through its modality

to the term it defines. As a simple example, consider the definition of anemia; in part, at least, the definition says that anemia is a condition in which erythrocytes are deficient in number. The lack of erythrocytes can be represented in a cognitive network by a proposition in which some property, say few, applies to the entity erythrocyte. But anemia is neither a fewness nor an erythrocyte; it is a condition. No part of the proposition defines the condition, but the whole proposition does--at least in part. Hence anemia is at a different level of abstraction from the terms that define it, and nothing in other versions of cognitive networks seems capable of expressing this distinction.

A slightly more complex example is necessary to show the power of metalingual definitions. Consider glucose-6-phosphate deficiency anemia. Like anemia, this condition is partially defined by deficiency of erythrocytes. But it is further characterized by a deficiency of glucose-6-phosphate. And the latter deficiency is causally related to the former. The metalingual definition of the condition therefore includes at least two propositions with their modalities, linked causally by a discursive arc, the whole subsumed under a modality as head.

Metalingual definitions can be more complex still; they can include spatial configurations, temporal sequences, relationships of purpose, and whatever else can appear in a

cognitive network. A definition can be as brief as a phrase, or as long as a textbook chapter.

Abstract or metalingual definition provides the only natural means of passage among the several languages of medicine. Each disease entity is ordinarily characterized in each of the three principal languages: A textbook account gives a clinical picture, a statement of laboratory findings relevant to the diagnosis, and--insofar as the state of the science allows--a pathophysiological explanation. These are three distinct metalingual definitions of the same entity; the theory connects them, and justifies giving the same name, i.e. the name of the disease, to the three pictures.

N. J. Y. Woodhouse has recently reviewed Paget's Disease of Bone. In his introductory sketch, he uses all three languages:

> Clinical picture. Affected bones are usually thickened, deformed, and the overlying skin feels hot. Bone pain is sometimes severe.
>
> Laboratory findings. Osteoclasts and osteoblasts are increased in number. Serum alkaline phosphatase and, in severe cases, acid phosphatase levels are elevated. Urine hydroxyproline excretion is raised and there is an increase in bone calcium turnover, as measured by isotopic methods.
>
> Pathophysiology. Bone resorption and formation

occur at rates far in excess of the normal adult skeleton. [The chapter contains further remarks on metabolism and possible etiology.]

At different times, the record of an individual case may contain various subsets of these observations, possibly linked to a diagnosis of Paget's disease. A universal code for medicine must therefore be capable of registering all three languages; when enlargement and deformation of bone is observed on physical examination, the observation has to be recorded in clinical language, and when confirmed by radiological examination the same observation has to be recorded in the language of laboratory findings. The two types of observations have different values, and must be kept separate. In obscure cases, the observations must be recorded as such, without diagnosis; later, the observations may lead to formulation of a list of possible diseases from which differential diagnosis is to be made. And finally a single diagnosis can result, to which some but not necessarily all observations contribute.

Three conclusions may be drawn. First, the metalingual connections in individual histories, between the various languages of observation and the language of diagnosis, are part of the record and need to be recorded in order to reconstruct the handling of a case. Second, these connections in individual histories are justified by reference to authority, the textbook writer or reviewer who

speaks from the fullest humanly attainable knowledge of the science; these authoritative connections may not be expressed in the code, but they underlie it. Third, the structure of medicine is exceedingly complex (surely no surprise), with linkages at many levels. Diseases are related to diseases, as sarcoma is an occasional complication of Paget's disease. And the definitions of diseases are linked, in as much detail as science permits, as Woodhouse suggests a linkage between serum alkaline phosphatase levels and bone resorption and formation.

Processes

The history of computation is a slow shift of tasks from brain to computer. Work is always accomplished by a division of labor; as yet we are willing to discard the results of even the simplest computations when human judgment considers them faulty. Every use of SNOMed is certain to divide the work somehow between medical and support personnel and computers; this section is not to be read as a plea that the computer be given more work. All I wish to do is show that certain tasks require certain tools. My plea, if any, is that more tools be converted from abstract to concrete. A concrete tool has the virtue of being subject to criticism and, therefore, improvement; it is more teachable, and provides a more definite standard for the measurement of performance.

Encoding. The object of encoding is to provide, at a specified level of depth, a complete record which may include diagnostic procedures, results, diagnostic inferences, orders, and therapeutic procedures. The encoding must be reliable, in the sense that different coders would write the same record in the same situation. It must be valid, in the sense that encoders and designers of subsequent procedures must agree on the significance of the records. Depth of encoding must vary across applications.

The tools suggested are a vocabulary, in systematic (paradigmatic) order and an index in alphabetical order. Since no coder will ever know the system completely, the coder needs help in finding correct encodings. The vocabulary serves this purpose, suggesting paradigmatically related terms if the coder can find approximately the right point of entry.

Some additional tools that would help are a file of generally valid propositions and a file of metalingual definitions. For example, entering the file of propositions under the heading bone would yield such propositions as "bone softens", "bone enlarges", and "bone deforms". Now, bone is not paradigmatically related to soften, enlarge, or deform, and so the vocabulary is of no help. The list of propositions can suggest appropriate terminology for the morphological changes to which bone is subject, the procedures that can be applied, and so on.

The existing files of metalingual definitions are textbooks, encyclopedias, and reviews. There may at present be no better form in which to present abstract definitions; but at least they have to be recognized as an adjunct to the code.

In attempting to devise computer programs for encoding, these files are necessary. Conversion from natural language to formalized language depends in part on grammar and in part on semantics. Files of generally valid propositions and metalingual definitions are part of semantic analysis: a statement that a particular patient's bones are softened is recognizable as semantically well formed just because the general proposition that bone softens is known. The statement that a certain pattern of symptoms leads to a given diagnosis is recognizable as semantically valid just because the pattern of symptoms fits the metalingual definition of the disease diagnosed.

Case management. The record of a case is maintained in ready accessibility for the life of the case, i.e., until recovery or death. Explicit recording of discursive and metalingual relationships has the effect of improving the day-to-day presentation of the record. Procedures and their outcomes can be grouped according to purpose. If the system includes a file of metalingual definitions, each tentative diagnosis can be presented with a record of evidence already collected and of procedures available to increase the

evidence. This use of the computer for diagnostic support is not new, but is probably too expensive, and above all too little understood as yet, to be installed routinely. Yet it can be foreseen. The authoritative files will eventually be available by telecommunication from regional or national centers, so that the hospital does not have to maintain them. What contribution the designers of SNOMed can make to better understanding of this problem is therefore important.

Administration. Discussion at the conference suggests that the determination and justification of charges is now too narrowly based. In particular, justification can arise from all of the languages of medicine, not from diagnoses alone.

Assessment. To evaluate the handling of a case, one must have a chronological account of procedures and inferences. Discursive relationships and metalingual definitions therefore provide a part of the necessary basis.

Research. Correlations among the clinical picture, the laboratory findings, and the pathophysiology are a traditional method of research in medicine. The value of systematically encoded files of medical records for this purpose has been asserted and, so far as I know, questioned only on the ground that the required files would be insupportably vast. The researcher needs, at least for some purposes, records written in a mixture of all three languages; however, he will often have to get his pathophysiology from other sources. The

responsibility of SNOMed's designers is to be as clear as possible about both the intended significance of each term, and the implicit or suppressed relations of terms to each other, so that the research worker can formulate his questions with a reasonable effort.

Modification. A code consists of a collection of entities and, possibly, a collection of relationships among them. With the relationships, the code becomes a cognitive network, whether the designers wish it so or not. For good reason, SNOMed is a cognitive network; the division of its terms into six fields, the hierarchical assignment of character strings, and the implicit relationships among the fields in a line of record give it structure.

A code without structure is modified without difficulty. As long as the number of items encodable in its format is not exceeded, it adds and deletes items freely.

A code with structure is modified with more difficulty. What is most difficult to alter is the implicit structure of the code. Once SNOMed is widely used, the transfer of a term from one field to another will be expensive. To shift an item from one paradigmatic position to another will render files, some of them perhaps large and of permanent value, obsolete. Science must change, and obsolescence of files is the inevitable concomitant, whether the files be in natural or formalized language, in brains, on paper, or in computers.

Conclusions

Any cognitive network needs to record paradigmatic, syntagmatic, discursive, and metalingual arcs; attitudinal arcs are sometimes needed, and certainly in medicine.

Medical record files must be able to keep all types of linkages to provide accurate histories, not only of the patient's condition but also of the course of diagnosis and treatment.

Medical authority files need all types of linkages to record the structure of the science.

Explicit structure imposes a burden on coders, who must think it out and write it down, and increases file size. Implicit structure imposes limitations on coders, who cannot always fit their observations or actions to the implicit structure, and on revisers, who make files obsolete by structural alteration. The balance is subtle; the problem merits the designers' closest attention.

Authority files and record (history) files can be separated. In that case, some relationships in the records must be recalculated from the authorities every time they are used. The necessity, certainly actual and probably permanent, of such separation goes to show that medical records will never be complete. If the code shows that tonsil is part of pharynx, and only the authority shows that tonsil is also part of lymphatic structure, it will be impossible in general to determine whether the surgeon who

removed the tonsils from his patient's pharynx knew that he was also removing them from his patient's lymphatic structure. But some points have to be taken for granted.

References

Rumelhart, David E., Peter H. Lindsay, and Donald A. Norman. A process model for long-term memory. In Organization of memory, edited by Endel Tulving and Wayne Donaldson. Academic Press, 1972, pp. 198-246. (This volume contains other relevant papers.)

Shank, Roger C. Conceptual dependency: a theory of natural language understanding. Cognitive psychology, vol. 3, no. 4, 1972, pp. 552-631.

Woodhouse, N. J. Y. Paget's disease of bone. In Clinics in endocrinology and metabolism, volume 1, number 1, edited by Iain MacIntyre. W. B. Saunders, 1972, pp. 125-141.